YOLOv5를 이용한 개인정보 탐지 및 마스킹 알고리즘 구현

유수연, 박유나, 서지혜, 오지연
덕성여자대학교 컴퓨터공학전공 학부생

dbtndus0527@gmail.com, yndbsk9372@gmail.com, wisdxx@gmail.com,
ohjy0530@gmail.com

Implementation of personal information detection and masking algorithm using YOLOv5

Su-Yeon Yoo, You-na Park, Ji-Hye Seo, Ji-Yeon Oh
Dept. of Computer Engineering, Duksung Women's University

요 약

미디어 공유 활성화에 따라 개인정보 노출의 위험성이 증가하고 있다. 본 논문에서는 YOLOv5를 통해 학습한 모델을 바탕으로 이미지에서 개인정보가 담긴 물체를 탐지하는 연구를 진행하였다. 모델로 탐지된 객체에는 광학 인식 기술(OCR)을 적용하여 객체 내의 텍스트 속 개인정보 영역을 마스킹한다. 설계된 알고리즘은 여러 분야에 활용되어 개인정보에 대한 서비스 강화를 제공해 줄 수 있을 것으로 기대된다.

1. 서론

개인정보 유출로 인한 금전적 피해 및 범죄 노출의 심각성은 우리에게 익히 알려져 있다. 하지만 많은 사람은 자신이 올리는 글, 이미지, 동영상 등에서의 개인정보 유출 민감도가 낮다. 따라서 개인이 콘텐츠를 올리기 전, 개인정보가 있는지 검사하고 그 개인정보를 마스킹할 필요가 있다.

본 논문에서 구축하는 개인정보 객체 탐지 및 마스킹 알고리즘은 YOLOv5를 사용하여 개인정보가 포함된 객체를 탐지한다. 그리고 이미지에서 텍스트를 추출하기 위해 개인정보가 포함된 객체에 광학 문자 인식(Optical Character Recognition, OCR) 기술을 적용하여 객체 내의 텍스트 영역을 인식한 뒤 해당 영역을 마스킹하고자 한다.

2. 개인정보 탐지 및 마스킹 모델 구축

2.1. 데이터 라벨링 및 데이터셋 구축

한국과학기술정보연구원 NTIS의 개인정보 영향도 등급[1]이 1등급에 해당하는 고유식별정보, 신용 및 금융정보와 2등급의 개인식별정보를 마스킹할 정보로 선정했다.

<표 1> NTIS 개인정보 영향도 등급

등급	분류	개인정보 항목
1등급	고유식별정보	주민등록번호, 여권번호, 운전면허번호, 외국인등록번호
	신용 및 금융정보	신용정보, 신용카드번호, 계좌번호 등
2등급	개인식별정보	이름, 주소, 전화번호, 핸드폰번호, 이메일 등
	개인관련정보	학력, 직업, 키, 몸무게, 가족상황 등
3등급	제한적 본인식별정보	회원정보, 사번 등

선정한 정보가 외부에 유출될 위험이 있는 객체를 데이터 수집 대상으로 포함한다. 해당 객체가 포함된 이미지를 직접 수집하고 Bounding Box로 이미지를 라벨링하여 모델이 분류할 수 있도록 데이터셋을 구축한다. 현재 이미지 내에 여권, 운전면허증, 신용카드, 주민등록증이 포함된 이미지 데이터를 수집했고 각각 passport, drivers-license, credit-card, national-id-card로 라벨링하여 객체를 분류한다. 수집한 데이터는 이미지 증강을 통해 약 22000개의 데이터를 확보하여 개인정보 객체 데이터셋을 구축했다.

2.2. YOLOv5 모델 규모

YOLOv5의 모델 규모로 "s", "m", "l", "x"가 있다. 본 논문에서는 YOLOv5의 모델 규모 "m"을 사용하였다. 모델 규모 "s"는 경량화되어 빠른 추론 속도를 제공하지만, 작은 객체에 대한 정확도가 제한될 수 있다. 모델 규모 "l"과 "x"는 높은 정확도를 제공하지만 보다 많은 컴퓨팅 리소스를 요구한다. 모델 규모 "m"은 정확도와 경량화의 균형을 잘 유지하고 있어 다양한 객체 크기와 높은 성능을 고려할 때 적절한 선택으로 보인다.

2.3. 하이퍼 파라미터 조정

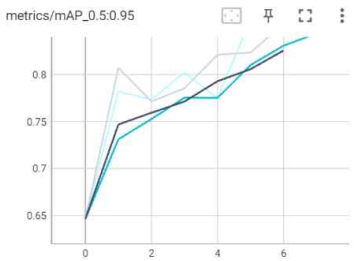
본 논문에서는 배치(batch) 크기와 에포크(epochs)를 조정하였다. 배치 크기는 GPU 메모리 용량과 모델 훈련 안정성을 고려하여 16으로 설정되었다. 에포크는 데이터셋의 특성과 학습 속도를 고려하여 50으로 설정하였다.

2.4. 모델 학습 및 테스트 결과

객체 탐지 모델의 정밀도와 일반화 능력을 평가하기 위해 평균 정밀도(mean Average Precision, mAP)를 주요 평가 지표로 선택하였다.

<표 2> 학습 횟수에 따른 mAP 값

	학습 횟수					
mAP	1	2	3	4	5	6
측정값	0.731	0.773	0.802	0.774	0.858	0.860



(그림 1) mAP 추이 그래프

객체 탐지 모델은 IoU 임계값 범위 0.5에서 0.95까지의 mAP 값으로 0.860을 기록하였다. 이 결과는 모델이 다양한 객체 크기와 위치에서 안정적으로 성능을 발휘함을 나타낸다.

초기에는 mAP 값이 낮게 시작하였으나, 학습이 진행됨에 따라 점차 증가하는 경향을 그래프에서 확인할 수 있다. 특히 mAP값이 0.731에서 0.860으로 증가한 것을 보아 모델의 성능이 향상되었음을 시각적으로 확인할 수 있다.

2.5 개인정보 텍스트 마스킹

광학 문자 인식(OCR)은 이미지의 텍스트를 기계가 읽을 수 있는 텍스트 데이터로 변환하는 기술이다. 본 논문에서는 탐지한 객체 영역 내에서 OpenCV를 통해 이미지 전처리를 진행한다. 전처리한 이미지에서 광학 문자 인식으로 텍스트 영역을 추출하고 개인정보에 해당하는 텍스트 영역을 마스킹한다.

3. 결론

인터넷을 통한 미디어 공유가 활성화됨에 따라 개인정보의 노출을 예방하고자 YOLOv5를 사용하여 개인정보 노출의 여부를 판단할 수 있는 개인정보 객체 탐지 및 마스킹 알고리즘을 구현했다.

개인정보가 포함된 이미지 데이터는 방대한 데이터를 구하기 어려워 한정된 이미지를 증강하여 학습을 진행하다 보니 현재 학습된 모델은 오버피팅을 보인다. 해당 문제를 보완하기 위해 2.1에서 언급한 4가지 항목에 대한 학습 데이터를 추가로 확보하여 모델을 학습할 예정이다. 또한, 학습할 객체를 우편번호, 차 번호판, 택배 송장 스티커 등으로 확장하여 다양한 객체를 인식할 수 있도록 모델을 학습할 예정이다. 이미지 내에서 텍스트 인식률이 낮아 텍스트 추출의 정확도가 낮은 상태이다. 추후 이미지 전처리 방법을 보완하여 인식률을 향상할 계획이다.

본 논문에서 구현한 개인정보 객체 탐지 및 마스킹 알고리즘은 SNS, 스트리밍 서비스 등에 적용될 수 있다. 해당 알고리즘이 다양한 서비스에 확장된다면, 개인정보 보안이 강화된 인터넷 환경뿐 아니라 개인정보 유출에 대한 사용자의 민감도 증가 효과를 가져올 것으로 기대된다.

Acknowledgement

본 프로젝트는 과학기술정보통신부 정보통신창의인재양성사업의 지원을 통해 수행한 ICT멘토링 프로젝트 결과물입니다.

참고문헌

[1] 박석천 “이미지 인식 기반 향상된 개인정보 식별 및 마스킹 시스템 설계 및 구현”, 한국인터넷방송통신학회 논문지, vol.17, no.5, pp. 1-8, 2017